Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 8973867, 9 pages
https://doi.org/10.1155/2018/8973867

Research Article

Constructing a High-Order Globally Convergent Iterative Method for Calculating the Matrix Sign Function

Haifa Bin Jebreen

Mathematics Department, College of Science, King Saud University, Riyadh, Saudi Arabia

Correspondence should be addressed to Haifa Bin Jebreen; hjebreen@ksu.edu.sa

Received 9 April 2018; Accepted 29 May 2018; Published 21 June 2018

Academic Editor: Alberto Cavallo

This work is concerned with the construction of a new matrix iteration in the form of an iterative method which is globally convergent for finding the sign of a square matrix having no eigenvalues on the axis of imaginary. Toward this goal, a new method is built via an application of a new four-step nonlinear equation solver on a particulate matrix equation. It is discussed that the proposed scheme has global convergence with eighth order of convergence. To illustrate the effectiveness of the theoretical results, several computational experiments are worked out.

1. Preliminaries

The sign function for the scalar case is expressed by

$$\operatorname{sign}(\omega) = \begin{cases} 1, & \operatorname{Re}(\omega) > 0, \\ -1, & \operatorname{Re}(\omega) < 0, \end{cases} \tag{1}$$

wherein $z \in \mathbb{C}$ is not located on the imaginary axis. Roberts in [1] for the first time extended this definition for matrices, which has several important applications in scientific computing, for example see [2–4] and the references cited therein. For example, the off-diagonal decay of the matrix function of sign is also a well-developed area of study in statistics and statistical physics [5].

To proceed formally, let us consider that $A \in \mathbb{C}^{n\times n}$ is a square matrix possessing no eigenvalues on the axis of imaginary. We consider

$$A = TJT^{-1}, \tag{2}$$

as a form of Jordan canonical written such that

$$J = \operatorname{diag}(J_1, J_2), \tag{3}$$

and the eigenvalues of $J_1 \in \mathbb{C}^{p\times p}$ are in the open left half-plane, while the eigenvalues of $J_2 \in \mathbb{C}^{q\times q}$ are in the open right half-plane. It is now possible to write the following [6]:

$$S = \operatorname{sign}(A) = T \begin{pmatrix} -I_p & 0 \\ 0 & I_q \end{pmatrix} T^{-1}, \tag{4}$$

wherein $p + q = n$. Noting that $\operatorname{sign}(A)$ is definable once A is nonsingular.

This procedure takes into account a clear application of the form of Jordan canonical and of the matrix T. Herein none of the matrices T and J are unique. However, it is possible to investigate that $\operatorname{sign}(A)$ as provided in (4) does not rely on the special selection of T or J.

Here, a simpler interpretation for the sign matrix in the case of Hermitian case (namely, all eigenvalues are real) can be given by

$$S = U\operatorname{diag}(\operatorname{sign}(\lambda_1), \ldots, \operatorname{sign}(\lambda_n))U^*, \tag{5}$$

wherein

$$U^*AU = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \tag{6}$$

is a diagonalization of the square matrix A.

The significance of calculating and finding S in (4) is because of the point that the function of sign plays a central role in matrix functions theory, specially for principal matrix

roots and the polar decomposition; for more one may refer to [7–9].

Bini et al. in [10] proved that the principal p-th root of the matrix A could be written as a multiple of the (2,1)-block associated with the sign matrix sign(C), associated with the block companion matrix:

$$C = \begin{pmatrix} 0 & I & & & \\ & 0 & I & & \\ & & \ddots & \ddots & \\ & & & \ddots & I \\ A & & & & 0 \end{pmatrix} \in \mathbb{C}^{pn\times pn}. \tag{7}$$

It is requisite to focus of the most general case, i.e., when all the eigenvalues are complex rather than being narrow over a range of specific matrices, such as in (5).

An important point goes to the fact that although sign(A) is a square root of the unit matrix, it is not equal to I or $-I$ unless the A's spectrum locates completely in the open right or left half-plane(s), respectively. Thus, the sign function is a nonprimary square root of I.

Apart from (4), an efficient way to derive matrix iterations for some matrix functions is to apply the zero-finding iterative methods for solving operator equations which here is a matrix equation. Toward this goal, it is necessary to tackle a nonlinear equation as comes next:

$$F(X) \coloneqq X^2 - I = 0, \tag{8}$$

wherein I is a unit matrix, so as to propose matrix methods for S. The main motivation in this work is to extend the recently published results of the literature [11, 12] in this category by providing a useful novel method for calculating sign matrix. Furthermore, the proposed procedure can be applied for the calculation of polar decomposition, principal matrix square root, and several other scientific computing problems.

After this brief introduction about the matrix function of sign in Section 1, the remaining sections of this study are given as comes next. Section 2 shortly surveys the existing matrix iterations and their importance for computing S. In Section 3, it is discussed how we construct a new method having global convergence behavior and not belonging to the class of Padé family of iterations for computing S. It too manifests that the proposed scheme is convergent with high order of convergence. Computational reports are furnished to illustrate the higher computational precision of the constructed solvers in Section 4. A final remark of the manuscript is given in Section 5 with some directions for future studies.

2. The Literature

In the current research work, iterative methods are the main focus for calculating S. As a matter of fact, such matrix iteration methods are Newton-type schemes that are in essence fixed-point schemes by providing a convergent matrix sequence by imposing a sharp initial value.

The connection of matrix iterative expressions with the function of sign is not that straightforward, but in practice, such methods could be constructed by considering a suitable root-finding method to the nonlinear matrix equation (8). Noting that sign(A) is a solution of this equation (refer to [12] and the references cited therein).

Applying the classic Newton's method (NM) to (8) yields

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1}\right). \tag{9}$$

An inversion-free version of (9), called Newton-Schultz iteration [6], is defined by

$$X_{k+1} = \frac{1}{2}X_k\left(3I - X_k^2\right), \tag{10}$$

by applying the well-known Schulz inverse-finder in order to remove the computation of the inverse matrix per computing step.

The Newton-Schulz scheme is a second-order convergent, inversion-free method in calculating the sign matrix, but it suffers from the drawback that its convergence unlike the Newton's method (9) is local.

Analogously, the third-order convergent Halley's method (HM) [13] for calculating the sign matrix is defined by

$$X_{k+1} = \left[I + 3X_k^2\right]\left[X_k\left(3I + X_k^2\right)\right]^{-1}. \tag{11}$$

It is noted that all the above-mentioned schemes are particular cases of the Padé family presented and discussed in [13, 14]. The Padé approximation belongs to a broader category of rational approximations. Coincidentally, the best uniform approximation of the sign function on a pair of symmetric but disjoint intervals can be expressed as a rational function.

Recently a fourth-order iterative method was furnished in [15] as follows:

$$X_{k+1} = \left[I + 18X_k^2 + 13X_k^4\right]\left[X_k\left(7I + 22X_k^2 + 3X_k^4\right)\right]^{-1}. \tag{12}$$

3. Construction of a New Matrix Method

Assume the following nonlinear equation:

$$g(x) = 0, \tag{13}$$

wherein $g : D \subseteq \mathbb{C} \to \mathbb{C}$ is a scalar function. In what follows, let us first present a new scheme in nonlinear equation solving. The idea of increasing the order (see, e.g., [16, 17]) is to consider several substeps, while the newly appearing first derivatives are approximated via a secant-like approximation. Thus, we may write

$$y_k = x_k - s_k,$$

$$z_k = x_k - \left(1 + \frac{g(y_k)}{g(x_k) - (5/3)\,g(y_k)}\right)\frac{g(x_k)}{g'(x_k)},$$

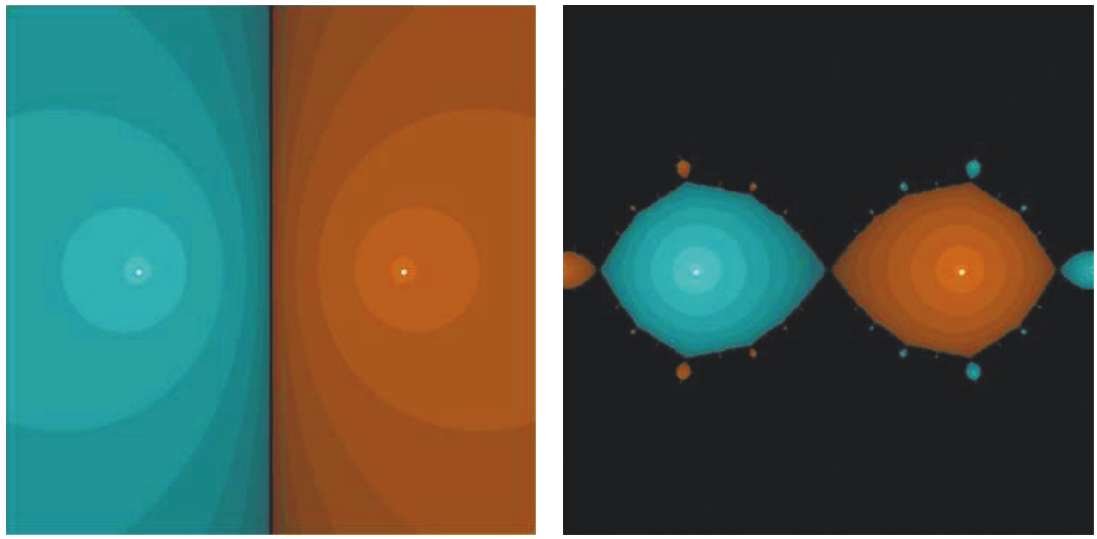

Figure 1: Basins of attractions for (9) in left and (10) in right (shading is done based on the number of cycles to achieve the convergence).

$$w_k = z_k - \frac{g(z_k)}{g[y_k, z_k]},$$
$$x_{k+1} = w_k - \frac{g(w_k)}{g[w_k, z_k]}, \tag{14}$$

whereas the first-order divided difference operator is defined by

$$g[\ell, \hbar] := \frac{g(\hbar) - g(\ell)}{\hbar - \ell}. \tag{15}$$

Here the point is that the new method should not aim at having the *optimal convergence order* (in the sense of Kung-Traub, see, e.g., [16]) since such schemes lose the global convergence behavior when applied to (13). Since the final objective is to propose a method for matrix sign, we should take two things into account, which are having global convergence behavior and the novelty. In fact, the optimality conjecture of Kung-Traub is not useful once we extend iterative methods for calculating the sign of a matrix. Because, the optimality conjecture here ruins the final structure of the matrix method.

One may also ask that why method (14) has been selected and what are the other ways for improving it. To answer these, we recall that the best way to improve the performance of (14) is to add one Newton's substep at the end of its fourth step, which is costly since it includes the computation of the first derivative per cycle. In a similar way as in (14), we can add a secant-like substep and increase the convergence order. Generally speaking, a family of iterations can now be constructed in this way. On the other hand, since very higher order methods may not be useful in double precision arithmetic, namely, the practical environment that most researchers work in, we here only provide (14) and discuss its application and extension for matrix sign.

Theorem 1. *Assume that $\alpha \in D$ is a simple root of a function $g : D \subseteq \mathbb{C} \to \mathbb{C}$, which is sufficiently differentiable and contains x_0 as an initial value. Accordingly, the iterative expression (14) reads*

$$e_{k+1} = \frac{1}{9}(c_2)^7 e_k^8 + \mathcal{O}(e_k^9), \tag{16}$$

where $c_j = g^{(j)}(\alpha)/j!g'(\alpha)$, and $e_k = x_k - \alpha$.

Proof. The steps of proving the convergence order for this iterative method are via Taylor expansion, which is straightforward. □

Applying (14) on the matrix equation (8) will yield a novel matrix scheme to calculate (4) in its reciprocal form as follows:

$$X_{k+1} = X_k\left(12I + 200X_k^2 + 560X_k^4 + 344X_k^6 + 36X_k^8\right)\Xi_k^{-1}, \tag{17}$$

wherein

$$\Xi_k = \left[I + 64X_k^2 + 406X_k^4 + 532X_k^6 + 145X_k^8 + 4X_k^{10}\right], \tag{18}$$

and the initial approximation is

$$X_0 = A. \tag{19}$$

Applying (14) on the nonlinear equation $g(x) = x^2 - 1$ contributes a global convergence in the complex plane (excluding the values locating on the axis of imaginary). The basins of attraction for (10) (locally convergent) and (9) (globally convergent) are portrayed in Figure 1. This global behavior of the proposed scheme, that is kept for matrix case, has been shown in Figures 2-3

To draw the basins of attractions, we consider a square $\Gamma = [-2, 2] \times [-2, 2] \in \mathbb{C}$ and allocate a color to any point

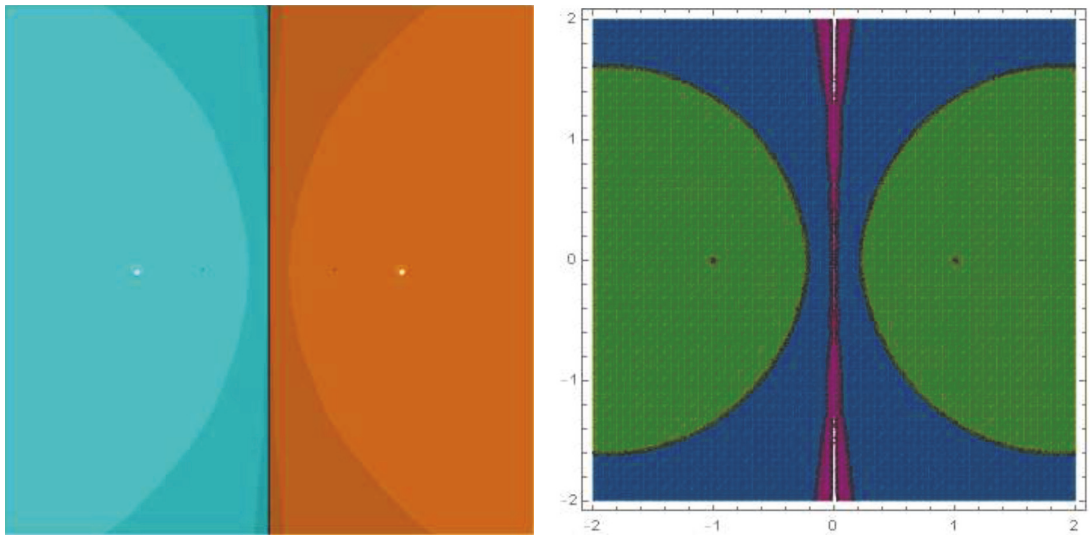

Figure 2: Basins of attractions for (17) in left and its density plot on the same domain in right (shading is done based on the number of cycles to achieve the convergence).

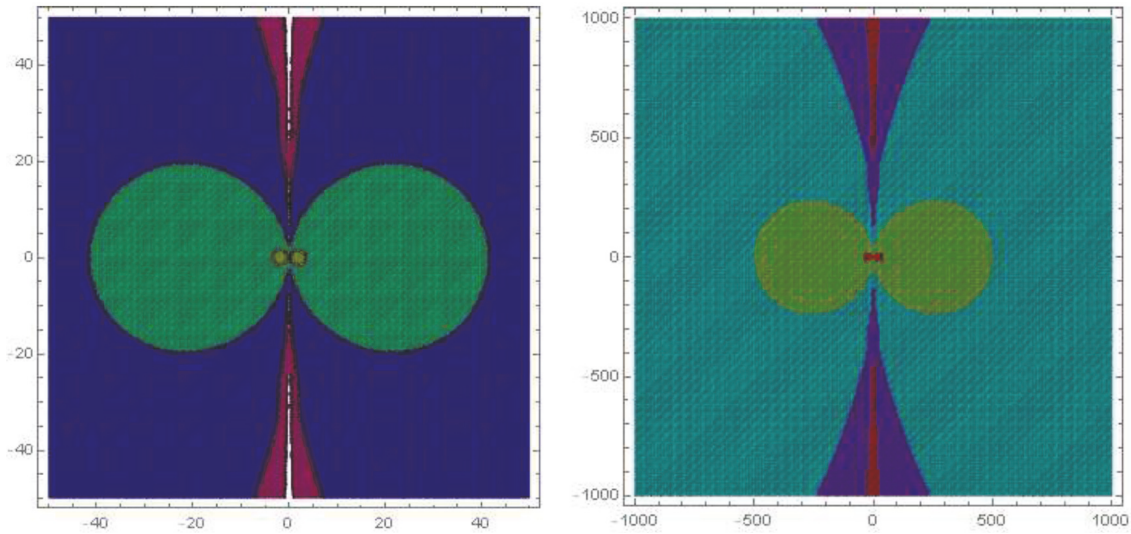

Figure 3: The density plot of the basins of attractions for (17) on $[-50, 50] \times [-50, 50]$ in left and on $[-100, 100] \times [-100, 100]$ in right (shading is done based on the number of cycles to achieve the convergence).

$x_0 \in \Gamma$ according to the simple zero, at which the new methods (or the existing methods for comparisons) converge. Subsequently, we highlight the point as black once the scheme diverge. Herein, we consider the stopping termination for convergence as $|g(x_k)| \leq 10^{-2}$. Using a different stopping criterion $|x_{k+1} - x_k| \leq 10^{-2}$, the density plot of the basins of attraction only for the new high order of convergence method (17) are brought forward in Figures 2-3 on different domains.

3.1. Convergence Study. Now, it is shown that the proposed schemes are convergent, under standard conditions, namely, when there are no pure imaginary eigenvalues in the absence of rounding errors.

Theorem 2. *Assume that $A \in \mathbb{C}^{n\times n}$ possess no eigenvalues on the axis of imaginary. Accordingly, the iterations $\{X_k\}_{k=0}^{k=\infty}$ expressed by (17) is convergent to S, using (19).*

Proof. Assume that R is a rational operator in accordance with (17). Since $X \in \mathbb{C}^{n\times n}$ has a form of Jordan canonical, there exists a matrix Z, so that

$$X = ZJZ^{-1}. \tag{20}$$

It is recalled that diagonalizable and nondiagonalizable matrices have a Jordan normal form $A = TJT^{-1}$, whereas J includes the Jordan blocks. So,

$$R(X) = ZR(J)Z^{-1}. \tag{21}$$

An eigenvalue λ of X_k is transferred into an eigenvalue of $R(\lambda)$ of X_{k+1} via the iterative expression (17). This relevance and relation among the eigenvalues show that it is required to search how $R(\lambda)$ transforms the complex plane into itself. It is recalled that R has the feature of sign preservation, namely,

$$\operatorname{sign}(R(x)) = \operatorname{sign}(x) \quad x \in \mathbb{C}. \tag{22}$$

Moreover, it should have the global convergence, that is, the sequence defined by

$$x_{k+1} = R(x_k), \tag{23}$$

with $x_0 = x$ converges to $\operatorname{sign}(x)$ while x is not located on the axis of imaginary. At this moment, assume that the square matrix A have a form of canonical Jordan considered just like [6, p. 107]:

$$Z^{-1}AZ = \Lambda = \begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}, \tag{24}$$

wherein Z is a not singular and C, N are square Jordan blocks in association with eigenvalues locating in $\mathbb{C}^-$ and $\mathbb{C}^+$, respectively. We show by $\lambda_1, \ldots, \lambda_p$ and $\lambda_{p+1}, \ldots, \lambda_n$ locating on the main diagonals of blocks C and N, respectively. Applying (24), it is possible to write

$$\operatorname{sign}(A) = Z \begin{bmatrix} -I_p & 0 \\ 0 & I_{n-p} \end{bmatrix} Z^{-1}. \tag{25}$$

Taking (25) into account, it is easy to deduce

$$\operatorname{sign}(\Lambda) = \operatorname{sign}\left(Z^{-1}AZ\right) = Z^{-1}\operatorname{sign}(A)\, Z = \begin{pmatrix} \operatorname{sign}(\lambda_1) & & & & & 0 \\ & \ddots & & & & \\ & & \operatorname{sign}(\lambda_p) & & & \\ & & & \operatorname{sign}(\lambda_{p+1}) & & \\ & & & & \ddots & \\ 0 & & & & & \operatorname{sign}(\lambda_n) \end{pmatrix}. \tag{26}$$

From $D_0 = Z^{-1}AZ$, we define

$$D_k = Z^{-1}X_kZ, \tag{27}$$

in order to obtain a convergent sequence to $\operatorname{sign}(\Lambda)$. Thence, from the scheme (17), we simply could obtain

$$\begin{aligned} D_{k+1} = {} & D_k\left(12I + 200D_k^2 + 560D_k^4 + 344D_k^6 + 36D_k^8\right) \\ & \times \left[I + 64D_k^2 + 406D_k^4 + 532D_k^6 + 145D_k^8 \right. \\ & \left. + 4D_k^{10}\right]^{-1}. \end{aligned} \tag{28}$$

When D_0 is a diagonal matrix, it is possible to show that all successive D_k are diagonal matrices, via mathematical induction. The other case when D_0 is not diagonal and will be handled in the remaining part of the proof.

By rearranging (28) as n uncoupled scalar iterations as comes next:

$$d_{k+1}^i = \frac{12d_k^i + 200{d_k^i}^3 + 560{d_k^i}^5 + 344{d_k^i}^7 + 36{d_k^i}^9}{1 + 64{d_k^i}^2 + 406{d_k^i}^4 + 532{d_k^i}^6 + 145{d_k^i}^8 + 4{d_k^i}^{10}}, \tag{29}$$

where

$$d_k^i = (D_k)_{i,i}, \quad 1 \le i \le n. \tag{30}$$

Using (28) and (29), we should investigate the convergence of $\{d_k^i\}$ to $\operatorname{sign}(\lambda_i)$, for all $1 \le i \le n$. From (29) and because the eigenvalues of A are not pure imaginary, it is possible to write

$$\operatorname{sign}(\lambda_i) = s_i = \pm 1. \tag{31}$$

Thus, we attain

$$\frac{d_{k+1}^i - s_i}{d_{k+1}^i + s_i} = -\left(\frac{-s_i + d_k^i}{s_i + d_k^i}\right)^8 \left(\frac{s_i - 2d_k^i}{s_i + 2d_k^i}\right)^2. \tag{32}$$

Noting that the factor $(s_i - 2d_k^i)/(s_i + 2d_k^i)$, is bounded due to choosing an appropriate initial matrix (19). Since

$$\left|\frac{d_0^i - s_i}{d_0^i + s_i}\right| < 1, \tag{33}$$

we attain

$$\lim_{k\to\infty}\left|\frac{d_{k+1}^i - s_i}{d_{k+1}^i + s_i}\right| = 0, \tag{34}$$

and, therefore,

$$\lim_{k\to\infty}\left(d_k^i\right) = s_i = \operatorname{sign}(\lambda_i). \tag{35}$$

At this point, it possible to derive that

$$\lim_{k\to\infty} D_k = \operatorname{sign}(\Lambda). \tag{36}$$

Noting that if D_0 is not diagonal, the relation among the iterates' eigenvalues must be dealt with for (17). The eigenvalues of X_k are transformed from the iteration k to the iteration $k+1$ via

$$\begin{aligned} \lambda_{k+1}^i = {} & \left(12\lambda_k^i + 200{\lambda_k^i}^3 + 560{\lambda_k^i}^5 + 344{\lambda_k^i}^7 + 36{\lambda_k^i}^9\right) \\ & \times \left[1 + 64{\lambda_k^i}^2 + 406{\lambda_k^i}^4 + 532{\lambda_k^i}^6 + 145{\lambda_k^i}^8 \right. \\ & \left. + 4{\lambda_k^i}^{10}\right]^{-1}, \quad 1 \le i \le n. \end{aligned} \tag{37}$$

The relation (37) manifests that the eigenvalues generally are convergent to $s_i = \pm 1$, namely,

$$\lim_{k\to\infty}\left|\frac{\lambda_{k+1}^i - s_i}{\lambda_{k+1}^i + s_i}\right| = 0. \tag{38}$$

Ultimately, it would be easy to write that

$$\lim_{k\to\infty} X_k = Z\left(\lim_{k\to\infty} D_k\right) Z^{-1} = Z\,\mathrm{sign}\,(\Lambda)\, Z^{-1} = \mathrm{sign}\,(A)\,. \tag{39}$$

This ends the convergence proof for (17) to calculate S. □

Now by considering (18) and the facts that X_k are rational functions of A, so, similar to A, commute with S, and $S^2 = I$, $S^{-1} = S$, $S^{2j} = I$, and $S^{2j+1} = S$, $j \geq 1$. It can be shown that the new scheme reads the following error inequality:

$$\left\|X_{k+1} - S\right\| \leq \left(\left\|\Xi_k^{-1}\right\| \left\|I - 2X_k\right\|^2\right) \left\|X_k - S\right\|^8. \tag{40}$$

Inequality (40) shows the eighth order of convergence.

A scaling technique to accelerate the initial phase of convergence is normally requisite since the convergence rate cannot be seen in the initial iterates. Such an idea was discussed fully in [18] for Newton's method. A robust procedure to improve the initial convergence speed is to scale the iterations before each iteration; i.e., X_k should be moved to $\mu_k X_k$.

If the scaling parameter (for the Newton's method) is defined by [18],

$$\zeta_k = \begin{cases} \sqrt{\dfrac{\left\|X_k^{-1}\right\|}{\left\|X_k\right\|}}, & (\text{norm scaling}), \\ \sqrt{\dfrac{\rho\left(X_k^{-1}\right)}{\rho\left(X_k\right)}}, & (\text{spectral scaling}), \\ \left|\det\left(X_k\right)\right|^{-1/n}, & (\text{determinantal scaling}), \end{cases} \tag{41}$$

then the accelerated forms of the proposed matrix iteration for S is defined as follows:

$$\begin{aligned} &X_0 = A, \\ &\zeta_k = \text{is the scaling parameter computed by (41)}, \\ &X_{k+1} = \zeta_k X_k \left(12I + 200\zeta_k^2 X_k^2 + 560\zeta_k^4 X_k^4 + 344\zeta_k^6 X_k^6 \right. \\ &\quad \left. + 36\zeta_k^8 X_k^8\right) \times \left[I + 64\zeta_k^2 X_k^2 + 406\zeta_k^4 X_k^4 + 532\zeta_k^6 X_k^6 \right. \\ &\quad \left. + 145\zeta_k^8 X_k^8 + 4\zeta_k^{10} X_k^{10}\right]^{-1}, \end{aligned} \tag{42}$$

where

$$\lim_{k\to\infty} \zeta_k = 1, \tag{43}$$

$$\lim_{k\to\infty} X_k = S. \tag{44}$$

4. Experiments

Herein, several experiments are discussed for the calculation of the sign matrix. The direct application of the new formulas for finding S is given below, though the application for computing the polar decomposition, finding the Yang-Baxter matrix equation can be given similarly. The simulations are run on an office laptop with Windows 7 Ultimate equipped Intel(R) Core(TM) i5-2430M CPU 315 2.40GHz processor and 16.00 GB of RAM on a 64-bit operating system. In addition, the simulations are done in Mathematica 11.0 [19].

Various schemes are compared in respect to the iteration numbers and the elapsed CPU time. Globally convergent schemes are only included for comparison. The compared matrix methods are NM, HM, ANM, and PM1 (i.e., (17)). We do not include comparisons with methods having local convergence behavior such as the Newton-Schulz method (10) or (computationally expensive) methods from different categories such as the ones based on the computation of the Cauchy integral

$$\mathrm{sign}\,(A) = S = \frac{2}{\pi}\int_0^\infty \left(t^2 I + A^2\right)^{-1} dt. \tag{45}$$

The stopping criterion for our simulations is defined by

$$\left\|X_k^2 - I\right\|_2 \leq 10^{-4}. \tag{46}$$

The reason of choosing (46) lies in the fact that, at each iterate, the obtained approximation should satisfy the main matrix equation. Thus, this criterion is much more trustful than other Cauchy-like terminations when calculating the sign of a matrix.

Experiment 1. Here, we calculate the sign matrix of 20 generated randomly *complex* matrices (with uniform distributions via the following piece of codes in the Mathematica environment) as comes next

```
SeedRandom[123]; number = 20;
Table[A[l] = RandomComplex[{-5 - 5 I, 5 + 5 I},{50 l, 50 l}];, {l,number}];
```

Noting that here $I = \sqrt{-1}$.

The numerical reports are provided in Tables 1-2 for various sizes of the input matrices based on the required number of steps and the elapsed CPU times. The results uphold the analytical parts and discussions of Sections 2-3. They manifest that there is a clear improvement in the iterations' numbers and the total elapsed CPU time by applying (17). As a matter of fact, the mean of number of iterations and the CPU times listed in the last rows of each

Table 1: Comparison of iterations' numbers for Experiment 1.

Matrix No.	NM	ANM	HM	PM1
$A_{50\times50}$	12	11	8	4
$A_{100\times100}$	15	14	10	5
$A_{150\times150}$	18	17	12	6
$A_{200\times200}$	15	14	10	5
$A_{250\times250}$	18	18	12	6
$A_{300\times300}$	23	21	14	6
$A_{350\times350}$	19	17	12	8
$A_{400\times400}$	17	16	11	6
$A_{450\times450}$	16	16	10	6
$A_{500\times500}$	20	20	13	7
$A_{550\times550}$	17	18	11	6
$A_{600\times600}$	20	19	13	6
$A_{650\times650}$	21	21	13	7
$A_{700\times700}$	17	18	11	6
$A_{750\times750}$	20	21	13	6
$A_{800\times800}$	20	22	13	6
$A_{850\times850}$	20	21	13	9
$A_{900\times900}$	20	21	13	8
$A_{950\times950}$	20	20	13	6
$A_{1000\times1000}$	21	24	14	8
Mean	18.45	18.45	11.95	**6.35**

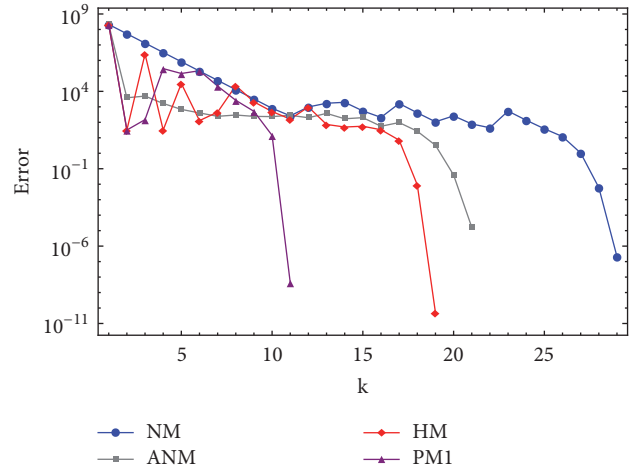

Figure 4: The convergence history of different methods in Experiment 2.

table indicate that the scheme (17) has the best performance in general.

It is pointed out that the calculation of X_k^2 per iteration for computing the stopping termination introduces one matrix product for NM, while the HM and the presented scheme calculate this matrix in the middle of each cycle.

Experiment 2. The aim of this test is to check the convergence history of different methods for a randomly 1000×1000 generated complex matrix as follows:

```
SeedRandom[1];
A= RandomComplex[{-100 - 100 I, 100 + 100 I}, {1000, 1000}];
```

using the stopping criterion

$$\text{Error} = \left\| X_k^2 - I \right\|_F. \tag{47}$$

The results for Experiment 2 are given in Figure 4 showing a stable and consistent behavior of the proposed scheme for finding matrix sign function.

The numerical reports and evidences in Section 4 improve the *mean* of the CPU time, clearly. This was the main target of this paper in order to propose an efficient method.

5. Discussion

In various fields of numerical linear algebra and scientific computing, the theory and computation of matrix functions are very much useful. In the modern numerical linear algebra, it underlies an effective way introducing one to resolve the topical problems of the control theory. The function of a matrix can be defined in several ways, of which the following three are generally the most useful: Jordan canonical form, polynomial interpolation, and finally Cauchy integral.

In this research work, we have focused on iterative methods for this purpose. Hence, a high-order nonlinear equation solver has been employed for constructing a novel scheme in calculating the sign matrix, which does not have pure imaginary eigenvalues.

It was shown that the convergence is global via attraction basins in the complex plane and the rate of convergence is eight. Finally, some numerical experiments in double precision arithmetic were performed to manifest the superiority and applicability of (17). Outlines for future works can be forced to extend the discussed matrix iterations for calculating polar decompositions in future studies based on the application of the new schemes.

Data Availability

All the data used for comparison of different methods in this article have been generated using random generators, via the programming package Mathematica. The data can be generated in this way. Moreover, interested readers may contact the corresponding authors if they need any of such programming codes for further studies.

Table 2: Comparison of the elapsed (CPU) times for Experiment 1.

Matrix No.	NM	ANM	HM	PM1
$A_{50\times50}$	0.148498	0.0230943	0.00861964	0.0102797
$A_{100\times100}$	0.0943384	0.109786	0.0727258	0.0392688
$A_{150\times150}$	0.232303	0.506882	0.189903	0.137378
$A_{200\times200}$	0.330597	0.709762	0.280812	0.21293
$A_{250\times250}$	0.657373	1.44232	0.555998	0.458002
$A_{300\times300}$	1.29408	2.56028	1.13594	0.801164
$A_{350\times350}$	1.60244	2.90168	1.36108	1.53705
$A_{400\times400}$	2.06766	3.83108	1.84136	1.8719
$A_{450\times450}$	2.6057	5.23433	2.41827	2.14952
$A_{500\times500}$	4.38739	9.09064	4.03174	3.20028
$A_{550\times550}$	4.89925	10.7225	4.83132	3.62766
$A_{600\times600}$	7.07534	14.5142	7.22801	4.66437
$A_{650\times650}$	9.72633	19.8393	8.51814	6.87333
$A_{700\times700}$	9.34249	21.6787	8.03064	7.05715
$A_{750\times750}$	13.9398	31.1319	11.5503	8.65203
$A_{800\times800}$	16.7687	39.0702	14.9441	10.4224
$A_{850\times850}$	19.9251	45.4667	18.0246	18.9427
$A_{900\times900}$	24.1888	52.8144	20.6969	19.2772
$A_{950\times950}$	28.187	58.5243	24.4635	17.3038
$A_{1000\times1000}$	34.3487	82.3412	29.5465	26.1742
Mean	9.09109	20.1257	7.98652	**6.67063**

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research project was supported by a grant from the "Research Center of the Female Scientific and Medical Colleges", Deanship of Scientific Research, King Saud University.

References

[1] J. D. Roberts, "Linear model reduction and solution of the algebraic Riccati equation by use of the sign function," *International Journal of Control*, vol. 32, no. 4, pp. 677–687, 1980.

[2] P. Benner and E. S. Quintana-Ort, "Solving stable generalized Lyapunov equations with the matrix sign function," *Numerical Algorithms*, vol. 20, no. 1, pp. 75–100, 1999.

[3] J. L. Howland, "The sign matrix and the separation of matrix eigenvalues," *Linear Algebra and its Applications*, vol. 49, pp. 221–232, 1983.

[4] C. S. Kenney and A. J. Laub, "The matrix sign function," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 40, no. 8, pp. 1330–1348, 1995.

[5] J. Hardin, S. R. Garcia, and D. Golan, "A method for generating realistic correlation matrices," *The Annals of Applied Statistics*, vol. 7, no. 3, pp. 1733–1762, 2013.

[6] N. J. Higham, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2008.

[7] Z. Bai and J. Demmel, "Using the matrix sign function to compute invariant subspaces," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 1, pp. 205–225, 1998.

[8] R. Byers, C. He, and V. Mehrmann, "The matrix sign function method and the computation of invariant subspaces," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 3, pp. 615–632, 1997.

[9] J. Leyva-Ramos, "A note on mode decoupling of linear time-invariant systems using the generalized sign matrix," *Applied Mathematics and Computation*, vol. 219, no. 22, pp. 10817–10821, 2013.

[10] D. A. Bini, N. J. Higham, and B. Meini, "Algorithms for the matrix pth root," *Numerical Algorithms*, vol. 39, no. 4, pp. 349–378, 2005.

[11] B. Laszkiewicz and K. Zietak, "Algorithms for the matrix sector function," *Electronic Transactions on Numerical Analysis*, vol. 31, pp. 358–383, 2008.

[12] F. Soleymani, P. S. Stanimirovic, S. Shateyi, and F. K. Haghani, "Approximating the matrix sign function using a novel iterative method," *Abstract and Applied Analysis*, vol. 2014, Article ID 105301, 9 pages, 2014.

[13] C. Kenney and A. J. Laub, "Rational iterative methods for the matrix sign function," *SIAM Journal on Matrix Analysis and Applications*, vol. 12, no. 2, pp. 273–291, 1991.

[14] O. Gomilko, F. Greco, and K. Zitak, "A Padé family of iterations for the matrix sign function and related problems," *Numerical Linear Algebra with Applications*, vol. 19, no. 3, pp. 585–605, 2012.

[15] F. Soleymani, E. Tohidi, S. Shateyi, and F. . Haghani, "Some matrix iterations for computing matrix sign function," *Journal of Applied Mathematics*, vol. 2014, Article ID 425654, 9 pages, 2014.

[16] T. Lotfi, K. Mahdiani, P. Bakhtiari, and F. Soleymani, "Constructing two-step iterative methods with and without memory," *Computational Mathematics and Mathematical Physics*, vol. 55, no. 2, pp. 183–193, 2015.

[17] F. K. Haghani and F. Soleymani, "An improved Schulz-type iterative method for matrix inversion with application," *Transactions of the Institute of Measurement and Control*, vol. 36, no. 8, pp. 983–991, 2014.

[18] C. Kenney and A. J. Laub, "On scaling Newton's method for polar decomposition and the matrix sign function," *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 3, pp. 698–706, 1992.

[19] M. L. Abell and J. P. Braselton, *Mathematica by Example*, Academic Press, Netherlands, 5th edition, 2017.